Statistical Analysis Plan

Interventional, randomised, double-blind, placebo-controlled, active reference (fluoxetine), fixed-dose study of vortioxetine in paediatric patients aged 7 to 11 years, with Major depressive disorder (MDD)

Vortioxetine

Study No.:	12709A
Sponsor:	H. Lundbeck A/S (Lundbeck) 2500 Valby (Copenhagen), Denmark
[redacted]	[redacted]
SAP edition No.:	2.0
SAP edition date:	8 February 2022
Edition history:	Edition 1.0 (original SAP), 30 November 2015 (the version No. in the footer is the system version No.)

Table of Contents

Appendices

List of Panels and Tables

List of Abbreviations and Definitions of Terms

ALT	alanine aminotransferase
ANCOVA	analysis of covariance
APES	All-patients-enrolled set
APRS	all-patients-randomised set
APTS	all-patients-treated set
AST	aspartate aminotransferase
ATC	anatomical therapeutic chemical
BMI	body mass index
BPI	Brief Psychosocial Intervention
BUN	blood urea nitrogen
C-CASA	Columbia Classification Algorithm of Suicide Assessment
CDRS-R	Children depression rating scale revised version
CGAS	Children's Global Assessment Scale
CGI-I	Clinical Global Impression – Global Improvement
CGI-S	Clinical Global Impression – Severity of Illness
CI	confidence interval
CRP	C-reactive protein
C-SSRS	Columbia-Suicide Severity Rating Scale
DSM-5™	Diagnostic and Statistical Manual of Mental Disorders, 5th edition
ECG	Electrocardiogram
eCRF	electronic case report form
FAS	full-analysis set
GBI	General Behavior Inventory
HDL	high density lipoprotein
HR	heart rate
IMP	investigational medicinal product
INR	international normalised ratio of prothrombin time
K-SADS-PL	Kiddie-Schedule for Affective Disorders and Schizofrenia for School-aged Children, Present and Lifetime version
LDH	lactate dehydrogenase
LDL	low density lipoprotein
LOCF	last observation carried forward
MAR	missing at random
MASC-10	Multidimensional Anxiety Scale for Children short version
MDD	Major depressive disorder
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	mixed model for repeated measurements
MNAR	missing not at random

OC	observed cases
PAERS	Paediatric Adverse Event Rating Scale
PBO	placebo
PCS	potentially clinically significant
PedsQL	Pediatric Quality of Life Inventory
PGA	Parent Global Assessment – Global Improvement
PQ-LES-Q	Paediatric Quality of Life Enjoyment and Satisfaction Questionnaire
PRO	patient-reported outcome
PYE	patient years of exposure
QRS	specific ECG interval describing ventricular depolarisation
QT	specific ECG interval describing ventricular depolarisation/repolarisation
QTcB	heart-rate corrected QT interval using Bazett's correction formula
QTcF	heart-rate corrected QT interval using Fridericia's correction formula
REML	restricted maximum likelihood
SAE	serious adverse event
SAS	statistical software package from the SAS® Institute
SD	standard deviation
SOC	system organ class
TEAE	treatment-emergent adverse event

1 Objectives

1.1 Primary Objective

To evaluate the efficacy of vortioxetine 10 mg/day and 20 mg/day versus placebo after 8 weeks of treatment of depressive symptoms in children with a DSM-5™ diagnosis of Major depressive disorder (MDD).

1.2 Secondary Objectives

To evaluate the efficacy of vortioxetine 10 mg/day and 20 mg/day versus placebo during the 8 weeks of treatment on:

- global clinical impression
- functionality
- health-related quality of life

To assess pharmacokinetics of vortioxetine in paediatric patients, aged 7 to 11 years using population pharmacokinetic approach

1.3 Exploratory Objective

To explore the efficacy of vortioxetine 10 mg/day and 20 mg/day versus placebo on co-morbid symptoms

1.4 Safety Objective(s)

To evaluate the safety and tolerability of vortioxetine 10 mg/day and 20 mg/day versus placebo in children with a DSM-5™ diagnosis of MDD

2 Study Design

This study has been designed in accordance with the Declaration of Helsinki.[1]

This is an interventional, multi-national, multi-site, randomised, two-period, single- and double-blind, parallel-group, placebo-controlled, active-reference (fluoxetine), fixed-dose study.

An overview of the study is presented in Panel 1.

The study was originally designed as a 4-arm study, with two doses of vortioxetine (10 and 20 mg), placebo, and fluoxetine (20 mg, included as an active reference). Due to recruitment difficulties, the study design was amended (*Protocol Amendment 3*).

An interim analysis to potentially stop the study early for either efficacy or futility was performed when 271 randomized patients had either completed or been withdrawn from the study. An independent statistical group performed the interim efficacy analysis using the analyses for the primary hypothesis.

As the study continued after the interim analysis, enrolment to the fluoxetine arm was stopped and the study continued as a 3-arm study.

The study consists of a screening period of 5-15 days and consists of two Phases (Phase A and Phase B) after this screening period, and a safety follow-up period.

Phase A is a single-blind 4-weeks period where patients (and parents) are blinded and receive placebo + Brief Psychosocial Intervention (BPI). In case of incomplete improvement to BPI/placebo then patients are randomised into Phase B.

Phase B is a double-blind 8-weeks period where patients and investigators are blinded. In this Phase all patients continue to receive BPI and in addition either placebo, vortioxetine 10 mg, vortioxetine 20 mg, or fluoxetine 20 mg (NB: fluoxetine only prior to interim analysis).

The study includes male and female patients, aged ≥7 and <12 years at Screening Visit with a diagnosis of MDD according to DSM-5^{TM} Criteria and a CDRS-R score ≥45 and a Clinical Global Impression – Severity of Illness (CGI-S) score ≥4 at the Screening Visit and the Baseline A Visit (Week 0).

Approximately 600 patients were planned for enrolment to the 4-week treatment with standardised brief psychosocial intervention (BPI) and placebo period (Phase A). At the end of Phase A, a total of 438 patients with incomplete improvement were planned to be randomised to the 8 week double-blind Phase B. A blinded sample size re- assessment was performed prior to the interim analysis to ensure that the study was adequately powered, which resulted in an increase in sample size to 539 randomised patients (see Section 15).

Prior to interim analysis (Study part 1), patients were randomized in a 1:1:1:1 ratio to vortioxetine 10 mg/day, vortioxetine 20 mg/day, fluoxetine 20 mg/day, or placebo. After interim analysis (Study part 2), patients were randomized in a 1:1:1 ratio to vortioxetine 10 mg/day, vortioxetine 20 mg/day, or placebo.

Incomplete improvement in depressive symptoms is defined as a <40% reduction from Baseline A on the Children depression rating scale revised version (CDRS-R), and with a CDRS-R ≥40 total score and confirmed by a Parent Global Assessment- Global Improvement (PGA) value of >2.

Patients who do not fulfil the criteria of incomplete improvement in Phase A at Week 3 will be withdrawn from the study before Week 4. Patients who do not fulfil the criteria of incomplete improvement at week 4 will be withdrawn from the study and not participate in the double-blind Phase B of the study. No specific study related follow-up is required for these patients. However, as rescue procedure, they may receive up to 4 outpatient visits to the study site for consultations over a two-month period.

Patients will receive 5 sessions of standardised BPI at Week 1, 2, 3, 5, and 8.

Patients randomised to the 10 mg/day vortioxetine group will receive a lower initial dose (5 mg/day for 2 days) prior to receiving 10 mg/day. Patients randomised to the 20 mg/day vortioxetine group will receive a lower initial dose (5 mg/day for 2 days followed by 10 mg/day for 2 days and 15 mg/day for 2 days) prior to receiving 20 mg/day. Patients randomised to the fluoxetine 20 mg/day group will receive a lower initial dose (10 mg/day for 6 days) prior to receiving 20 mg/day.

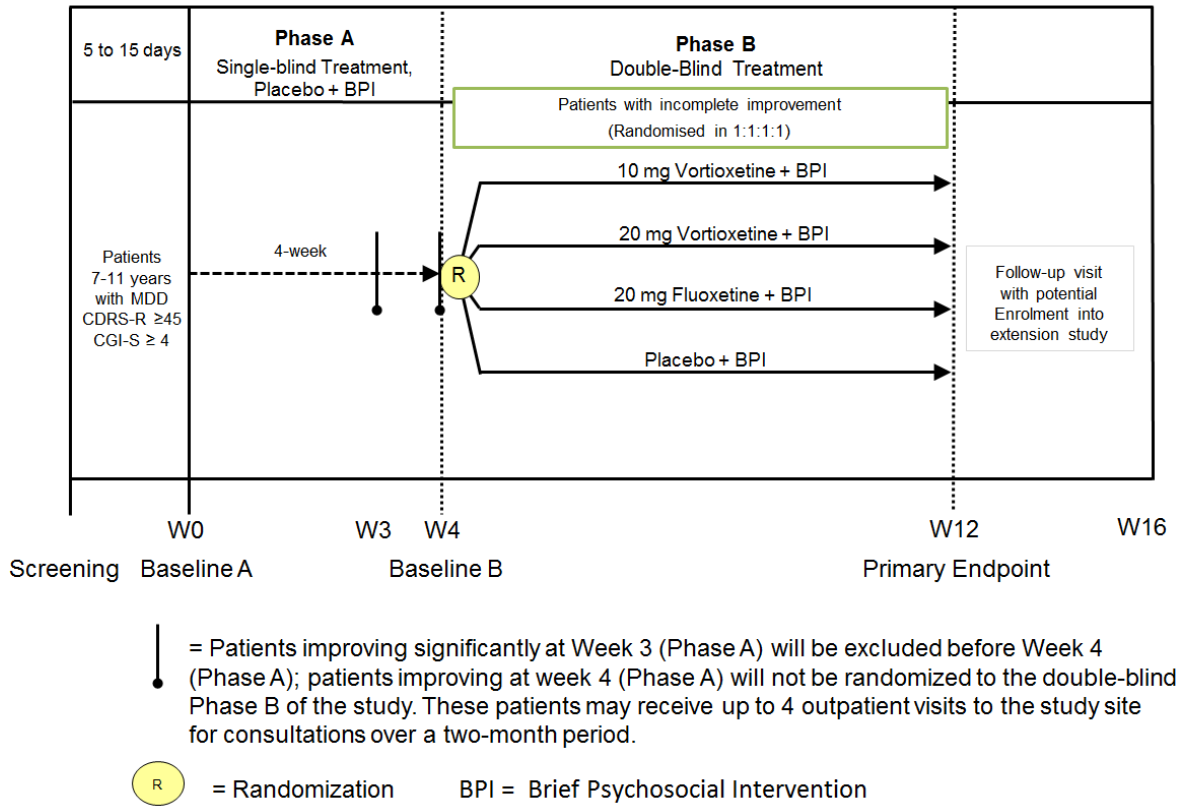
Based on tolerability the dose may be reduced at Week 8 (Visit 10 = Week 4 of Phase B). For patients randomised to vortioxetine the reduction will be by 5 mg/day and for fluoxetine 10 mg/day. No dose increase will be allowed.

An independent Data Monitoring Committee is established in order to review the safety and tolerability data throughout the study.

Panel 1 Study Design

Study part 1:

Applicable prior to interim analysis

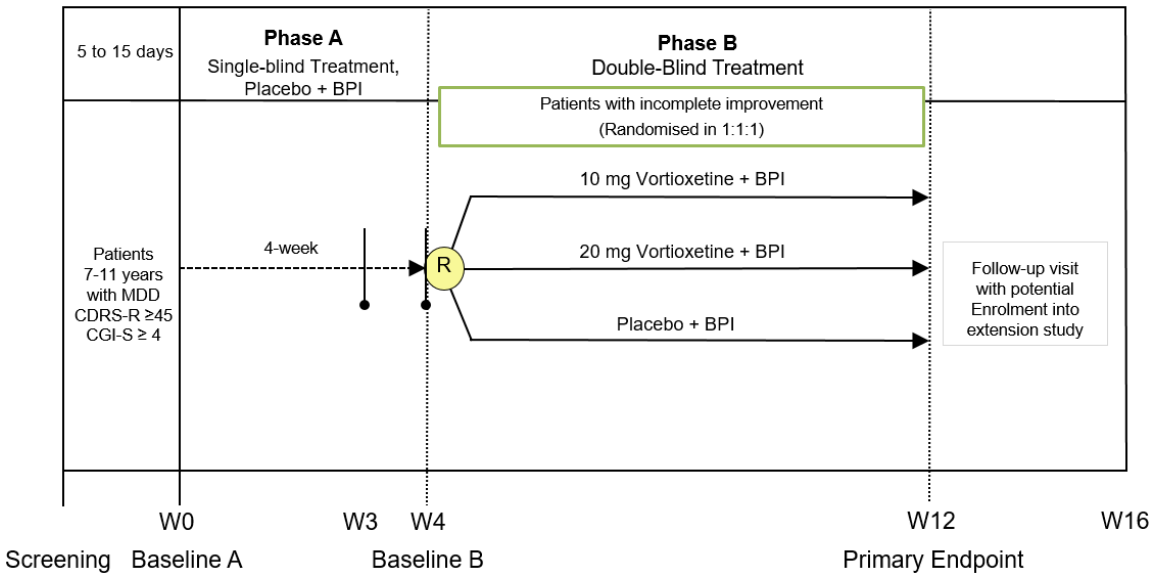


= Patients improving significantly at Week 3 (Phase A) will be excluded before Week 4 (Phase A); patients improving at week 4 (Phase A) will not be randomized to the double-blind Phase B of the study. These patients may receive up to 4 outpatient visits to the study site for consultations over a two-month period.

R = Randomization BPI = Brief Psychosocial Intervention

Study part 2:

Applicable after interim analysis (from June 14th 2019)



= Patients improving significantly at Week 3 (Phase A) will be excluded before Week 4 (Phase A); patients improving at week 4 (Phase A) will not be randomized to the double-blind Phase B of the study. These patients may receive up to 4 outpatient visits to the study site for consultations over a two-month period.

R = Randomization BPI = Brief Psychosocial Intervention

3 Endpoints

3.1 Primary Endpoint

Depressive symptoms:

- change from Baseline B in the CDRS-R total score after 8 weeks of treatment

3.2 Secondary Endpoints

Depressive symptoms:

- change from Baseline B in the CDRS-R total score during the 8 weeks of treatment
- change from Baseline B in the CDRS-R sub scores Mood (4 items), Somatic (6 items), Subjective (4 items), and Behaviour (3 items) during the 8 weeks of treatment
- CDRS-R response (defined as ≥50% reduction (calculated as change from baseline divided by (baseline value-17)) in the CDRS-R total score from Baseline B) during the 8-week treatment period
- remission over the 8-week treatment period (defined as CDRS-R ≤28), at each visit assessed

- change from Baseline B in the General Behaviour Inventory (GBI) depression total score, using the 10-item depression subscale, assessed by parent (PGBI-10D) and child (CGBI-10D), during the 8 weeks of treatment
- score in the PGA from one week after Baseline B during the 8-week treatment period

Global Clinical Impression:
- change from Baseline B in the CGI-S score during the 8-week treatment period
- score in the Clinical Global Impression – Global Improvement (CGI-I) from one week after Baseline B during the 8-week treatment period
- remission in the CGI-S score (defined as a CGI-S score of 1 or 2) during the 8 week treatment period, at each visit assessed

Functionality:
- change from Baseline B in the Children's Global Assessment Scale (CGAS) score during the 8-week treatment period
- change from Baseline B in the Pediatric Quality of Life Inventory (PedsQL) VAS score in each of the 6 domains during the 8-week treatment period
- change from Baseline B in the PedsQL average score over the 6 domains during the 8-week treatment period
- change from Baseline B in the PedsQL emotional distress summary score during the 8-week treatment period

Health-related quality of life:
- change from Baseline B in the Paediatric Quality of Life Enjoyment and Satisfaction Questionnaire (PQ-LES-Q) total score of item 1-14 during the 8 weeks treatment period
- change from Baseline B in the PQ-LES-Q overall evaluation score (item 15) during the 8-week treatment period

Pharmacokinetics:
- pharmacokinetic parameters for vortioxetine and fluoxetine

All pharmacokinetic analyses will be addressed in a separate PK analysis plan.

3.3 Exploratory Endpoint

Depressive symptoms:
- change from Baseline B in each CDRS-R item scores during the 8 weeks of treatment

Co-morbid symptoms:
- change from Baseline B in the Multidimensional Anxiety Scale for Children short version (MASC-10) total score during the 8-week treatment period

3.4 Safety Endpoints
- adverse events (AEs)

- Paediatric Adverse Event Rating Scale (PAERS)
- absolute values and changes from baseline in clinical safety laboratory tests, vital signs, weight, height and Electrocardiogram (ECG) parameters
- potentially clinically significant clinical safety laboratory test values, vital signs, weight, and ECG parameter values
- C-SSRS categorisation (suicidal ideation/suicidal behaviour)
- change from Baseline B in the General Behaviour Inventory (GBI) mania total score, using the 10-item mania subscale, assessed by parent (PGBI-10M) and child (CGBI-10M) during the 8 weeks of treatment

4 Analysis Sets

The following analysis sets will be used to analyse and present the data:

- *All-patients-enrolled* (APES) - all patients enrolled to the 4-week Single-blind period (Phase A)
- *all-patients-treated set Phase A* (APTS_A) – all patients in APES who received at least one dose of investigational medicinal product (IMP)
- *all-patients-randomised set Phase B* (APRS) – all patients randomised to the double-blind, 8-week treatment period (Phase B)
- *all-patients-treated set Phase B* (APTS) – all patients randomised to the double-blind, 8-week treatment period (Phase B) who took at least one dose of double-blind IMP
- *full-analysis set Phase B* (FAS) – all patients in the APTS who had a valid Baseline B assessment and at least one valid post-Baseline B assessment of the CDRS-R total score.

The patients and data will be classified into the analysis sets during a *Classification Meeting* according to the definitions above after the study database has been released but before the blind has been broken for the final analysis. The classification of the patients in Study part 1 should be kept at the final classification (after the end of the study).

5 Descriptive Statistics

Unless otherwise specified, summary statistics (n, arithmetic mean, standard deviation [SD], median, lower and upper quartiles, minimum and maximum values) will be presented for continuous variables, and counts and, if relevant, percentages will be presented for categorical variables.

Unless otherwise specified, data listings will include site, study Phase, treatment group, patient screening number, sex, age, race, and baseline A weight.

6 Patient Disposition

6.1 Summary of Patient Disposition

Patient disposition will be summarised by study Phase and by treatment group (in Phase B only) and include the number of patients in each analysis set defined in chapter 4, and the number of patients in Phase A (based on APES) and in Phase B (based on APRS) who completed or withdrew from the study.

6.2 Withdrawal

The number of patients who withdrew from study will be summarised in Phase A and by treatment group in Phase B. A summary of primary reason for withdrawal will be prepared for Phase A and summarised by treatment group in Phase B. A summary of all reasons for withdrawal will be prepared for Phase A and by treatment group in Phase B.

In addition, as the study continued after the interim analysis, similar summaries will be created based on patients randomised before and after the drop of the fluoxetine arm, in order to check the potential introduction of bias in patient selection by implementing the major *Protocol Amendment 3.*

Patients who withdrew from study in each study Phase will be listed and the listings will include the number of days in the study until withdrawal, the number of days on IMP, the primary reason for withdrawal, and all reasons for withdrawal.

Kaplan-Meier plots of time to withdrawal from study for any reason will be presented by treatment group in Phase B. The time will be calculated from the date of first dose of IMP in the Phase B to the date of completion or withdrawal from study. Patients who completed the study Phase B will be regarded as censored.

All tables, graphs, and listings will be based on the APTS_A and APTS depending on the study Phase.

All tables and graphs will be repeated for the FAS for Phase B if relevant.

7 Demographics and Other Baseline Characteristics

Demographics (sex, age, and race), baseline characteristics (height, weight, and BMI), disease characteristics (MDE history: MDD history, including number of MDEs, year and duration of current episode, treatment type for MDD, and outcome/response and tolerability), Tanner score, and baseline efficacy variables, based on FAS, will be summarised in Phase A and by treatment group in Phase B using the relevant baselines. In addition, as the study continued after the interim analysis, selected summaries will be created based on patients randomised before and after the drop of the fluoxetine arm, in order to check the potential introduction of bias in patient selection by implementing the major *Protocol Amendment 3.*

The medical, neurological, psychiatric, and social histories as well as the concurrent medical, neurological, and psychiatric disorders will be coded using the Medical Dictionary for Regulatory Activities (MedDRA) and summarised.

A medical, neurological, or psychiatric history is a disorder that ended prior to the Screening Visit. A concurrent medical, neurological, or psychiatric disorder is a disorder that is ongoing at the Screening Visit.

Demographics, selected disease characteristics, and baseline characteristics will be summarised based on the APTS_A and the APTS depending on the study Phase.

8 Recent and Concomitant Medication

Recent and concomitant medication will be coded using the WHO Drug Dictionary (WHO-DDE).

For Phase A medications will be classified according to the start and stop time for the period and summarised by anatomical therapeutic chemical (ATC) code and generic drug name, and treatment group:

- medication discontinued prior to first dose of IMP
- concomitant medication continued after first dose of IMP
- concomitant medication started at or after first dose of IMP and before first dose of IMP in Phase B

For Phase B medications will be classified according to the start and stop time for the period and summarised by ATC code, and generic drug name, and treatment group:

- medication discontinued prior to first dose of IMP in Phase B for randomized patients
- concomitant medication continued after first dose of IMP in Phase B
- concomitant medication started at or after first dose of IMP in Phase B

The tables will be based on the APTS_A and APTS unless otherwise mentioned.

For assigning medications to an analysis phase, please see Data Handling Plan.

9 Exposure and Compliance

Exposure to IMP will be defined as:

date of last dose of IMP – date of first dose of IMP + 1

Exposure to IMP will be summarised by treatment group in Phase B using descriptive statistics, and include the patient years of exposure (PYE). PYE will be calculated as the sum of the number of days of exposure to IMP for each patient, divided by 365.25 days.

In addition, exposure to IMP will be categorised in intervals for Phase A (1 – 7 days, 8 – 14 days, 15 - 21 days, 22 – days) and for Phase B (1 - 7 days, 8 – 14 days, 15 – 21 days, 22 – 28 days, 29 - 42 days, 43- days) and summarised by treatment group (only Phase B) in each study Phase. Missing values of exposure will not be imputed.

The number of patients who had at least one dose reduction due to poor tolerability will be summarised by treatment group.

Non-compliance days are days on which no IMP has been taken, less than the full dose of IMP has been taken, or more than the full dose of IMP has been taken.

Compliance with IMP for an interval between two visits will be defined as:

{date of end of visit interval – date of beginning of visit interval + 1 – number of days of non-compliance in the visit interval} / {date of end of visit interval – date of beginning of visit interval + 1} × 100%

The visit interval will start on the day after the first visit of the interval and end on the day of the last visit, to reflect the interval in which compliance is captured.

Compliance with IMP for the entire study will be defined as the compliance for the interval between the Baseline A Visit and the Completion/Withdrawal Visit:

{date of Completion/Withdrawal Visit – date of Baseline A Visit + 1 – number of days of non-compliance between Baseline A Visit and Completion/Withdrawal Visit} / {date of Completion/Withdrawal Visit – date of Baseline A Visit + 1} × 100%

Compliance with IMP for Phase A will be defined as the compliance for the interval between the Baseline A Visit and the Baseline B Visit:

{date of Baseline B Visit – date of Baseline A Visit + 1 – number of days of non-compliance between Baseline A Visit and Baseline B Visit} / {date of Baseline B Visit – date of Baseline A Visit + 1} × 100%

Compliance with IMP for Phase B will be defined as the compliance for the interval between the Baseline B Visit and the Completion/Withdrawal Visit:

{date of Completion/Withdrawal Visit – date of Baseline B Visit + 1 – number of days of non-compliance between Baseline B Visit and Completion/Withdrawal Visit} / {date of Completion/Withdrawal Visit – date of Baseline B Visit + 1} × 100%

Compliance with IMP will be categorised as "≤80% compliant" or ">80% compliant". The number and percentage of patients in each category will be summarised in Phase A and by treatment group in Phase B, both by visit interval and for the entire study.

Exposure and compliance will be summarised based on the APTS_A for Phase A and APTS for Phase B, and repeated for the FAS for Phase B.

10 Efficacy

10.1 General Efficacy Analysis Methodology

Unless otherwise specified, all the efficacy analyses will be based on the FAS, and all efficacy analyses will be done on both averaged and individual doses on vortioxetine.

As randomization to the fluoxetine arm has not been a possibility after the interim analyses, the number of fluoxetine patients in the final analyses will be substantially lower than for the placebo and vortioxetine patients.

All the tables and graphs will be presented by treatment group in Phase B.

Unless otherwise stated**,** all the p-values will be based on two-sided tests; the confidence intervals (CIs) will be two-sided. P-values will be displayed with 4 decimal points in all relevant outputs.

For endpoints not included in multiplicity adjustment, nominal p-values will be presented together with nominal 95% CIs.

All efficacy assessments will be summarised using descriptive statistics by visit for Phase A based on the APTS_A and by treatment group in Phase B based on the APTS.

10.2 Interim Analysis

An interim analysis was conducted in July 2019.

The interim analysis was based on a sequential approach (see *Interim Statistical Analysis Plan*), including an interim analysis with binding stopping rules for efficacy and futility, and an error-spending approach based on Kim & DeMets method with rho=2 were applied on the outcome from the MMRM model.[2]

Details of the monitoring boundaries for the sequential approach used at the interim analysis are provided in the *Interim Statistical Analysis Plan*.

An overview of the analysis flow from interim to final analysis is shown in section 10.4.1 and the SAS programs to be used in Appendix IV.

10.3 Testing Strategy

The testing strategy is applied on the model in section 10.4.1 and is as follows with an overall significance level for the study of 2.5% one-sided:

Step 1: Test the average dose effects of the two vortioxetine doses against placebo at final analysis based on a one-sided test on the level obtained for the final analyses taking the alpha-

spending for the interim into account. This new alpha-level will be calculated after unblinding.

Step 2: If significance is achieved in step 1, the two pairwise comparisons of the of the two vortioxetine doses (i.e., 10 mg/day or 20 mg/day) versus placebo will be performed separately based on a one-sided test on the same alpha-level as above.

The reason why both tests in step 2 can be performed without further multiplicity adjustment follows from a formal closed testing argument.

A formal descriptions of the closed testing principle can be seen in Appendix III.

10.4 Analysis Methodology for the Primary Endpoint

10.4.1 Primary Analysis of the Primary Endpoint

Change from Baseline B in the CDRS-R total in Phase B, will be analysed using a restricted maximum likelihood (REML)-based Mixed Model Repeated Measurements (MMRM) approach with freely varying mean and covariance structure and with country, treatment (vortioxetine 10 mg/day, vortioxetine 20 mg/day, fluoxetine, and placebo), and Week as fixed factors and Baseline B CDRS-R total score as a continuous covariate, the treatment-by-week interaction, and Baseline B CDRS-R-by-Week interaction. The Kenward-Roger approximation will be used to estimate denominator degrees of freedom. The analysis will be based on the missing-at-random (MAR) assumption and performed using all available observations (observed cases [OC] data) in Phase B.

The testing strategy is described in section 10.3, where the comparisons will be based on the least squares means for the treatment-by-visit interaction effect. The primary comparison will be the average effect with weights 0.5 and 0.5 of the two vortioxetine doses compared to placebo at Week 8 in Phase B based on the SAS LSMESTIMATE statement. The testing strategy also includes the comparisons of the individual vortioxetine doses to placebo.

The performed interim analysis is taken into account using the following stepwise approach:

The results from the interim analysis were used to update the interim-boundaries, based on the actual information fraction obtained at that time point.

These updated interim-boundaries are then after unblinding combined with the results from the MMRM analysis to obtain the final alpha-level used in the evaluation of the primary analysis.

This final alpha-level will then also be used for analysing the two separate vortioxetine doses versus placebo.

The details of these steps can be seen in the SAS code for the primary analysis in Appendix IV.

Both 95% CIs and (1-new alphalevel)*100% will be presented.

The country contributing the largest number of patients will be specified as reference group in the SAS CLASS statement country (ref="XXX"), with XXX as the corresponding country code in order to ensure convergence. Should non-convergence appear despite this, the following sequence of steps will be used, continuing to the next step only if non-convergence persists:

1. Grouped country will be used in the model instead of country (see section 18.4)
2. Maximum likelihood (ML) method will be used instead of the restricted maximum likelihood (REML) method
3. If non-convergence still persists, the sequence of steps (REML – ML) will be repeated with the following different covariance structures until convergence is reached:
   a. First-order Ante Dependence
   b. Heterogenous Toeplitz
   c. Heterogenous Compound Symmetry
   d. Compound Symmetry

10.4.2 Rationale for Selected Analysis Method for the Primary Endpoint

The MMRM analysis uses all available data measured repeatedly over time and allows for evaluation of the treatment-by-time interaction. The MMRM analysis provides an unbiased estimate of the treatment effect under the assumption that missing data are MAR. The use of an analysis assuming missing not at random (MNAR) is more suitable as a sensitivity analysis to address the impact of clinically plausible deviations from MAR.[3],[4]

Published data support the robustness of the MMRM analysis regarding protection against type I error and against bias, also in situations with a non-negligible proportion of missing data. Using extensive simulations, it has been demonstrated that the type I error is only affected to a limited extent and that the bias is small under the assumption that 1/3 of the missing data are missing-not-at-random, even when there is a severe imbalance between the treatment groups in the proportion of withdrawals.[5]

The chosen MMRM follows logically from the design of the study. The explanatory part of the model consists of the variables country, baseline CDRS-R score, visit and treatment. In this study stratification by site is done, and therefore including site as an explanatory variable is recommended in guidelines.[6] In this case, country is used instead as the number of sites are very high. If needed, small countries may be grouped; the strategy for grouping countries is described in Section 18.4. The baseline CDRS-R total score is included in the model to account for any imbalances in the CDRS-R total score between treatment groups at baseline and because of the importance of baseline as a predictor of the endpoint. This is in accordance with the recommendations in guidelines.[6] The baseline CDRS-R total score by-visit interaction is included in the model to allow for different dependence of the baseline CDRS-R score at different visits, and the treatment-by-visit interaction allows for estimation of the treatment effect at all visit, based on all available data.

The unstructured covariance is feasible in this study because of the relative low number of visits, and allows for maximum flexibility in estimation. An unstructured modelling of time and the treatment-by-time interaction provides an assumption-free approach, does not require estimation of an inordinate number of parameters, and can be depended upon to yield a useful result - attributes well suited to the primary analysis.[7]

In case of non-convergence the hierarchy of the chosen covariance structures was made so incremental model assumptions was added, meaning that the numbers of degrees of freedom was chosen as conservatively as possible.

10.4.3 Model Assumptions for Analysis of the Primary Endpoint

The assumption of normality will be investigated on an exploratory basis by inspection of a QQ-plot of the residuals.

The assumption of homoscedastic residuals will be investigated on an exploratory basis by inspection of a scatter-plot of the residuals versus the fitted values and by treatment group.

10.4.4 Sensitivity Analyses of the Primary Endpoint

Sensitivity analyses will be performed to evaluate how different assumptions affect the estimates of the treatment effect.

Sensitivity analysis using a pattern mixture model (PMM) for missing (valid) values, where it is assumed that the trajectory of patients with missing values in the vortioxetine groups is the same as that of the patients in the placebo group. The model will impute an outcome almost certainly worse than that assumed by MAR if vortioxetine is more efficacious than placebo. Thus, this approach is likely to provide a conservative estimate of the treatment effect if there is a beneficial treatment effect for vortioxetine since it both penalises high withdrawal rates as well as higher withdrawal rates in the vortioxetine groups compared to placebo. Monotone missing values (missing values after last observed valid value) will be imputed using a sequential regression-based multiple imputation method, based on the imputation models established from the placebo group.[8]

The procedure for the PMM method will include the following steps:

1. To prepare data for the PMM, non-monotone missing values (missing values before last valid observation), will be imputed using a Monte Carlo Markov Chain (MCMC) methodology assuming non-monotone missing values are missing at random. The imputation will be done by treatment group, and the model will include Baseline B CDRS-R total score and change from baseline B in CDRS-R total score at week 2, 4, 6, and 8 (Phase B). The SAS® procedure SAS® PROC MI will be used, using seed=1952146, and 200 imputations (nimpute=200).
2. Perform PMM on the monotone missing data created in step 1, where the monotone missing values are assumed to follow a MNAR pattern. The distribution for patients in the vortioxetine treatment groups at time t with last observation at time t-1, will be assumed to be equal to the conditional distribution for the placebo group with the

corresponding past. The regression model will include CDRS-R total score at Baseline B and changes from Baseline B in CDRS-R total scores, by using SAS® PROC MI using seed=195467 and the MONOTON REG () option.

3. Assemble a dataset containing data for all patients, including the imputed data from time t to serve as predictors for the imputation of missing data at the next week
4. Repeat steps 2) to 3) sequentially over all post-randomization visits at weeks (t+1, t+2, …)
5. Each of the 200 complete datasets will be analysed using the same MMRM model as described in section 10.4.1.
6. The estimated treatment effects for average effect of the two vortioxetine doses and the individual treatment groups obtained for the imputed datasets in step 5 will be combined to produce a unique point estimate and confidence interval using Rubin's rule to form a unique point estimate and standard error (SE), taking into account the uncertainty of the imputation,[9] using SAS® PROC MIANALYZE.

A sensitivity analysis of the primary endpoint will also be performed using an analysis of covariance (ANCOVA) model by visit using both the last observation carried forward (LOCF) and observed data (OC), including country and treatment (placebo, vortioxetine 10 mg/day, vortioxetine 20 mg/day and fluoxetine 20 mg/day) as factors, and baseline B CDRS-R Total Score as a continuous covariate. The mean difference between the averaged doses of vortioxetine and placebo will be estimated from the model at Week 8 in Phase B (Study Week 12). The estimates will be presented with p-values and 95% CIs.

The influence of other covariates such as age, sex, weight, height, and puberty as a covariate will be investigated with an ANCOVA with treatment and country as factors and baseline B score as a covariate and with both the use of appropriate interactions and main terms. The influence of quadratic terms of the covariates might be investigated similarly. All of this may lead to detailed subgroup analysis.

Evaluation of impact of potential non-compliance by PK concentration will be done (see the popPK plan for the study): after removing patients where PK indicates that they have not taken IMP, assessed through estimating oral clearance (CL/F; patients with CL/F values above 120 L/h are regarded as non-compliant) and/or plasma concentrations below lower quantification limit (LLOQ) ,an analysis based on an MMRM model similar to the one used in the primary analysis will be conducted.

If relevant, the potential impact of the major *Protocol Amendment 3*, will be investigated with an MMRM similar to the model for the primary analysis, where a factor will be added to the model indicating whether patients were randomised before or after the implementation of the amendment.

A supplementary analysis evaluating potential influential sites will be performed using the primary analysis model. Either via exclusion of sites or by site analysis as appropriate.

The SAS code for the sensitivity analysis of the primary endpoint can be seen in Appendix IV.

10.5 Analysis of the Secondary Endpoints

Panel 2 Secondary Efficacy Endpoints

Endpoint	Description	Type
Change from Baseline B in the CDRS-R score during the 8 weeks of treatment	At each visit assessed after Baseline B	1
Change from Baseline B in the CDRS sub scores Mood, Somatic, Subjective, and Behaviour during the 8 weeks of treatment	At each visit assessed after Baseline B	1
CDRS-R response during the 8 weeks treatment period	At each visit assessed after Baseline B; defined as ≥50% reduction (calculated as change from baseline divided by (baseline-17)) in the CDRS-R total score from Baseline B) during the 8 weeks treatment period	2
Remission over the 8 week treatment period, at each visit assessed	defined as CDRS-R <28	2
Change from Baseline B in the GBI depression total score, both parent and child versions, using the 10-item depression subscale, during the 8 weeks of treatment	At each visit assessed after Baseline B	1
Change from Baseline B in the GBI mania total score, both parent and child versions, using the 10-item mania subscale, during the 8 weeks of treatment**	At each visit assessed after Baseline B	1
Score in PGA from two weeks after baseline B during the 8 weeks treatment period	At each visit assessed after Baseline B	1
Change from Baseline B in the CGI-S score during the 8 weeks treatment period	At each visit assessed after Baseline B	1
Score in the CGI-I from one week after Baseline B during the 8 weeks treatment period	At each visit assessed after Baseline B	1
Remission in the CGI-S score during the 8 weeks treatment period, at each visit assessed	At each visit assessed after Baseline B; defined as a CGI-S score of 1 or 2	2
Change from Baseline B in the CGAS score during the 8 weeks treatment period	At each visit assessed after Baseline B	1
Change from Baseline B in the PedsQL VAS score in the 6 domains during the 8 weeks treatment period	VAS score for each domain separately at each visit assessed after Baseline B	1
Change from Baseline B in the PedsQL average score over the 6 domains during the 8 weeks treatment period	At each visit assessed after Baseline B	1

Change from Baseline B in the PedsQL emotional distress summary score during the 8 weeks treatment period	At each visit assessed after Baseline B; defined as the average score of domains anxiety, sadness, anger, and worry	1
Change from Baseline B in the PQ-LES-Q total score during the 8 weeks treatment period	At each visit assessed after Baseline B	1
Change from Baseline B in the PQ-LES-Q overall evaluation scores during the 8 weeks treatment period	At each visit assessed after Baseline B	1

1 = continuous; 2 = binary;
**: this is a safety endpoint treated like an efficacy endpoint with a continuous outcome

Continuous secondary endpoints in Panel 2, type 1, will be analysed with an MMRM similar to the model specified for the primary endpoint with comparisons from the same model used for all time points. In addition**,** ANCOVA (OC and LOCF) will be performed per visit with treatment and country as factors and baseline B score as a covariate.

For analyses of the CGI-I score and PGA score, the CGI-I score and PGA score at Baseline B will be used as covariate.

For dichotomous outcomes, such as response and remission in Panel 2, type 2, the primary methodology for analysis at each week during Phase B (FAS, LOCF) will be logistic regression with treatment as factor and baseline B CDRS-R Total Score as a covariate, where the p-values and confidence intervals will be based on a Wald calculation. This will be supplemented by a similar analysis based on OC. In additional sensitivity analyses, patients having a missing value at the week analysed will be classified as a non-responder/non-remitter (NRI), and the same logistic regression will be applied for both classifications.

For dichotomous outcomes the average effect of the two vortioxetine doses will be based on a SAS LSMESTIMATE statement using SAS PROC LOGISTIC.

All secondary analyses will present results for both the average effect of vortioxetine against placebo as well as the individual doses of vortioxetine and fluoxetine against placebo.

All comparisons will be presented with 95% CI and p-values.

10.6 Analysis of the Exploratory Endpoint

Comparisons between the averaged vortioxetine doses and placebo, as well as 10 and 20 mg/day vortioxetine and placebo**,** for change from Baseline B in the CDRS-R item scores to week 8 in Phase B will be made using estimates from a MMRM model similar to the model specified for the primary endpoint.

Comparisons between the averaged vortioxetine doses and placebo, as well as 10 and 20 mg/day vortioxetine and placebo, for change from Baseline B in the MASC-10 total score to week 8 in Phase B will be made using estimates from a MMRM model similar to the model

specified for the primary endpoint. In addition, ANCOVA (OC and LOCF) will be performed with treatment and country as factors and baseline score as a covariate.

11 Safety

11.1 Adverse Events

11.1.1 General Methodology for Adverse Events

Unless otherwise specified, tables, graphs, and listings will be based on the APTS_A and the APTS depending on the study Phase.

Tables by preferred term and tables by system organ class (SOC) and preferred term will be sorted in descending order by the percentage of patients in the highest dose group of vortioxetine and then by the percentage of patients in lowest dose group of vortioxetine then by the percentage of patients in the fluoxetine 20 mg/day group and the placebo group.

Unless otherwise specified, the summaries of adverse events will include the total number and percentage of patients with an adverse event, and the total number of events.

In summaries of adverse events presented by intensity, the maximum intensity of the adverse event will be used for patients who have more than one intensity of that event within each study Phase. Adverse events for which information on intensity is missing will be classified as *severe*.

Listings of adverse events will be sorted by patient, study Phase, site, treatment group, screening number, and adverse event start date and include preferred term, investigator term, adverse event start Phase and date, days since first dose of IMP in the study Phase, duration of the adverse event, action taken, causality, intensity, seriousness, and outcome. For adverse events that change in intensity, each intensity will be included.

11.1.2 Handling of Adverse Events

11.1.2.1 Adverse Events Changing in Intensity or Seriousness

In ADaM data, changes in intensity are included as additional records (rows) to the originally reported adverse event. For instance, an adverse event that changes from mild to moderate will have two records, one with intensity mild and one with intensity moderate.

If an adverse event changes from non-serious to serious, in addition to the originally reported adverse event start date, the start date for the seriousness is reported. In ADaM data, adverse events that change from non-serious to serious will be included as two records, one with seriousness non-serious and one with seriousness serious.

11.1.2.2 Coding of Adverse Events

Adverse events will be coded using MedDRA version 21.0 or later.

11.1.2.3 Assigning Treatment Emergent Status to Adverse Events Records

Treatment Emergent status will be assigned to adverse event records based on the time of onset of the adverse event or worsening of the adverse event (increase in intensity compared to the preceding intensity or change from non-serious to serious) compared to first dose of IMP (for handling of change in intensity or seriousness are described in section 11.1.2.1, and handling of incomplete dates are described in section 18.3.4):

Treatment-emergent adverse event (TEAE) records -

- All records for an adverse event that starts at or after the first dose of single-blind IMP
- The record for the first worsening and all records that follow the first worsening for an adverse event that starts before first single-blind IMP, and where the worsening occurs at or after the first IMP

Adverse event are considered causally related to the IMP when the causality assessment by the investigator is *probable* or *possible*.

For further details, see Data Handling Plan.

11.1.2.4 Assigning Adverse Event Records to an Analysis Phase

All records for an adverse event will be assigned to an analysis phase. The analysis phase will determine in which analysis phase an adverse event record is reported, and an adverse event may be reported in more than one analysis phase but only if it worsens.

TEAE records will be assigned to an analysis phase according to handling of incomplete dates and are described in 18.3.4.

For further details, see Data Handling Plan.

11.1.3 All Adverse Events

All adverse events will be listed for the APES and for the APRS separately.

Overview of the PYE, numbers, and percentages of patients with TEAEs, serious adverse events (SAEs), adverse events leading to withdrawal, or who died will be provided based on the APTS_A and the APTS. For TEAEs, SAEs, and adverse events leading to withdrawal, the total number of events will be included.

11.1.4 Non-TEAE Records

Non-TEAE adverse events (pre-treatment adverse events) will be summarised by preferred term based on the APES.

11.1.5 Treatment-emergent Adverse Events

The following summaries will be presented in Phase A and by treatment group in Phase B:

- TEAEs by SOC and preferred term
- TEAEs by preferred term
- TEAEs with an incidence >=2% and >=5% in any treatment group by preferred term
- causally related TEAEs by SOC and preferred term
- TEAEs by intensity (mild/moderate/severe) and preferred term
- causally related TEAEs by intensity and preferred term

11.1.6 Deaths

All adverse events for patients who died will be listed for the APES.

11.1.7 Serious Adverse Events

All SAE records will be listed for Phase A (APTS_A) and Phase B (APRS) separately. For patients not entering the planned extension study new SAEs will be collected during the safety/follow up visit and summarised.

Treatment-emergent SAE records will be summarised in Phase A and by treatment group in Phase B:

- SOC and preferred term
- preferred term

11.1.8 Adverse Events Leading to Withdrawal

All adverse events leading to withdrawal will be listed for Phase A (APTS_A) and Phase B (APRS) separately.

TEAEs leading to withdrawal will be summarised in Phase A and by treatment group in Phase B:

- SOC and preferred term
- preferred term

11.1.9 Adverse Events Related to Tolerability

Incidence and prevalence rates in every two-week interval during Phase B will be presented for nausea based on the APTS for events with complete start and stop dates.

Incidence will be defined as the number of patients with a new event in the specified time interval. The denominator in each interval will be the number of patients without the event at the start of the interval plus the number of patients with an ongoing event at the start of the interval for which the event stops before the end of the interval.

Prevalence will be defined as the number of patients with the event at the start of the interval or who have the event during the interval. The denominator will be the number of patients at the start of the interval.

11.1.10 Adverse Events of Special Interest

Treatment-emergent nausea (preferred term), vomiting (preferred term) and insomnia (defined in Panel 3) will be summarised by treatment group in Phase B.

Panel 3 Insomnia related adverse events searches

Preferred term
Hyposomnia
Initial Insomnia
Insomnia
Middle Insomnia
Poor Quality Sleep
Sleep Disorder
Terminal Insomnia
Dyssomnia

11.2 General Methodology for Other Safety Data

Unless otherwise specified, tables, graphs, and listings will be based on the APTS_A and the APTS by treatment group.

Baseline will refer to either Baseline A or Baseline B depending on the study Phase.

The denominators for the summaries of a given variable will be based on the number of patients with non-missing values at a given visit or during the assessment period.

Descriptive statistics for the safety variables, both absolute values and changes from baseline in Phase B, will be presented by visit and the last assessment. All available assessments will be included in the identification of the last assessment (scheduled, re-assessments, and unscheduled).

For variables for which potentially clinically significant (PCS) values are defined based on a relative change from baseline summary statistics for the relative change from baseline will be presented. For children (7-17 years of age) other PCS values for diastolic and systolic blood pressure have been used than those for adults.

The number and percentage of patients with at least one PCS value at any post-baseline assessment time point will be summarised by variable. All available assessments will be included in the evaluation of PCS values (scheduled, re-assessments, and unscheduled).

For patients with post-baseline PCS values, listings will be provided including all available values for the variable, with flagging of PCS values and out-of-reference-range values.

The PCS definitions will be provided by H. Lundbeck and are stated in Appendix V.

11.3 Clinical Safety Laboratory Test Data

11.3.1 Data Presentation

All the clinical safety laboratory test values will be presented in conventional and/or Système International (SI) units.

Shift tables for all clinical lab parameters displaying shifts from normal to out-of-the-reference range from Baseline B to withdrawal/completion will be provided by Phase B treatment and include the numbers and percentages of patients.

11.3.2 Urinalysis

For urine dipsticks, for which the results are categorical values (for example, negative, trace, 1+, 2+), the number and percentage of patients will be summarised for each test by visit.

The microscopy results will be listed for patients with findings by assessment time point.

11.3.3 Potential Drug-induced Liver Injury (DILI)

Summaries will be presented by treatment group for Phase B.

Signals of DILI will be assessed according to the FDA guideline[10] using the following criteria:

- alanine aminotransferase (ALT) or AST >2×-, >3×-, 5×-, 10×-, or 20×ULN
- total bilirubin (BILI) >2×ULN
- alkaline phosphatase (ALP) >1.5×ULN
- ALT or AST >3×ULN AND BILI >1.5× or >2×ULN

In addition, assessment time points for patients for whom Hy's Law is potentially fulfilled will also be flagged in the listing (pHYL):

- ALT or AST >3×ULN AND
- BILI >2xULN AND
- ALP<2xULN

In the summaries, each patient should be counted only once using the maximum assessment, or the most severe for the combined criteria.

Patients fulfilling any of the individual criteria in Phase B (ALT/AST, ALP, or BILI) will be listed, and the listing will include all available ALT, AST, BILI, and ALP, BILI, GGT, and EOSEL values (absolute and normalised), sorted by assessment date in ascending order.

Evaluation of potential Drug-Induced Serious Hepatotoxicity (eDISH) will also be done by plots. Scatter plots of maximum ALT/AST versus maximum BILI will be presented for Phase B. The criteria for the individual tests will be considered separately (this means that the maximum of ALT/AST and the maximum BILI may not occur at the same assessment timepoint). The values will be normalised by the ULN (unit xULN) and the X-and Y-axes will be on the log scale. The plot will include a reference line for ALT/AST values >3xULN, and a reference line for BILI values>2xULN. Four quadrants are defined by the reference lines, where the right upper quadrant being the most specific indicator for a drug's potential for causing serious liver injury (Hy's law quadrant). The plot will include number of patients in each quadrant for each treatment group.

Subject line plots with values-by-time for ALT, AST, ALP, BILI, GGT and EOSLE (overlaid in the same plot) will be generated for patients with ALT/AST > 3xULN in Phase B. The test values will be normalised by the ULN (unit xULN) and the Y-axis will be on the log scale. All assessments at Visit 6 or in Phase B will be included, and the time will be days since first IMP in Phase B. Reference lines for the day of first-and last IMP in Phase B will be included. If there is more than one assessment at the same time point for a test, the maximum value will be used.

For further details, see Data Handling Plan.

11.3.4 Changes in Fasting and non-fasting Lipid and Glucose Concentrations

Shift tables displaying the change in classification for fasting and non-fasting lipids and fasting and non-fasting glucose from Baseline B to any visit in Phase B will be provided for each test and include the numbers and percentages of patients.

11.4 Vital Signs, Height, and Weight

11.4.1 Data Presentation

Descriptive statistics for the vital signs parameters in Phase A and Phase B will be presented by day and the last assessment in the Period.

11.5 ECGs

11.5.1 Data Presentation

Shift tables displaying shifts from normal to out-of-the-reference range ECG findings from Baseline B to the Withdrawal/Completion Visit will be provided and include the numbers and percentages of patients.

11.6 Other Safety Endpoints

11.6.1 Columbia-Suicide Severity Rating Scale (C-SSRS)

The C-SSRS was assessed:

- for lifetime (using the Baseline Version) – the C-SSRS assessment obtained at screening/baseline A that collects a lifetime recall
- for the past 12 months (using the Baseline Version) – the C-SSRS assessment obtained at screening/baseline A that collects a recall of the past 12 months
- at baseline (using the Since Last Visit Version) – the C-SSRS assessment obtained at baseline A that collects information for a pre-specified, limited period, for example, from screening to baseline
- post-baseline A (using the Since Last Visit Version) – the C-SSRS assessments obtained after baseline A

Note, all assessments will be regarded as valid for C-SSRS (that is, the criteria that the assessment date-date of last IMP must be ≤7 days in Phase B will not be applied).

For lifetime, past 12 months, Baseline A, and Phase A and Phase B, patients will be counted as having *No suicidal ideation or behavior* (patients that answered no to all items in Panel 4), or for the most severe suicidal ideation or behavior (severity is given by the ascending order of the items in Panel 4). Number and percentage of patients in each category will be presented. In the tables, number and percentage of patients with any non-suicidal self-injury behavior will also be included.

Lifetime, past 12 months, Baseline A will be presented for both APTS_A and APTS. Phase A and Phase B will be summarised for APTS_A, and APTS, respectively.

Panel 4 C-SSRS Scores

C-SSRS Score		Related to:
1	Wish to be dead	Suicidal ideation
2	Non-specific active suicidal thoughts	
3	Active suicidal ideation with any methods (not plan) without intent to act	
4	Active suicidal ideation with some intent to act, without specific plan	
5	Active suicidal ideation with specific plan and intent	
6	Preparatory acts or behaviour	Suicidal behaviour
7	Aborted attempt	
8	Interrupted attempt	
9	Non-fatal suicide attempt	
10	Completed suicide (only applicable for the post-baseline assessments)	

For patients with any post-baseline A suicidal behaviour (see Panel 4), listings will be prepared including all original C-SSRS items; C-SSRS scores related to suicidal behaviour will be flagged. The listing will be based on APES.

11.6.2 Paediatric Adverse Event Rating Scale (PAERS)

Note, all assessments will be regarded as valid for PAERS (that is, the criteria that assessment date-date of last IMP must be ≤7 days in Phase B will not be applied).

For each item, a severity score will be derived where the analysis value is defined as:

- If *Recently Present=No*, the analysis value will be *Absent*.
  Otherwise, the analysis value will be the reported Severity (*Mild*, *Moderate*, *Severe*, or *Extreme*)

The assessments for the PAERS will be prepared according to the description in the Data Handling Plan and all assessments will be assigned to an analysis phase, please see Data Handling Plan. For each item 1 to 43, the following parameters will be derived for each analysis phase following a baseline (all assessments within a phase will be included in the evaluation):

- *Worst severity score*:
  Worst severity score of the item (analysis value will be the worst severity)
- *Worsening of severity score compared to baseline*:
  Worsening of the item compared to the baseline (analysis value will be 'Yes' or 'No')

If a patient had a worsening of an item in a phase (Phase A or Phase B), all original assessments at the baseline (Baseline A for Phase A and Baseline B for Phase B) and in the phase will be listed.

For phase A, tables will be based on APTS_A, and listing will be based on APES.
For phase B, tables will be based on APTS, and listing will be based on APRS.

Also graphical presentations of the PAERS will be done.

Item 44 and 45 will only be listed, based on APES.

For further details, see Data Handling Plan.

11.6.3 General Behavior Inventory (GBI) 10-Item Mania Scale

The GBI total score and change from baseline will be summarised by visit in Phase A and by visit and treatment group in Phase B.

12 Remote visits

Some patients may have remote visits, for example due to the COVID-19 virus, and primary efficacy analyses and safety tabulations will be done with respect to this subpopulation of patients (see Data Handling Plan).

Patients with remote visits will be summarized and listed.

13 Pharmacokinetic Analyses

A separate analysis plan for pharmacokinetic analyses will be prepared by the Department of Quantitative Pharmacology, H. Lundbeck A/S.

14 Interim Analyses

An interim analysis for efficacy and futility has been performed (see section 10.2).

BIOSTATA was contracted to perform the analyses for the Data Monitoring Committee (DMC).

15 Sample Size Considerations

To obtain a power of 85%, with a one-sided significance level of 2.5% and an expected effect of 4 for each Vortioxetine dose, and a standard deviation of 11 for the change from baseline in CDRS-R for each dose, 102 patients needed to be included in a non-sequential trial.

To maintain the power at 85%, the sample size needed to be increased by a factor of 1.045 (see *Blinded sample size re-assessment*) to correct for the loss of power due to the sequential approach. This correction factor was obtained from R, and depends on the error spending function, its parameters, and the type I and type II errors.

As a result, 102*1.045≈107 randomised patients per arm (vortioxetine 10 mg/day, vortioxetine 20 mg/day, placebo) were required in the final analysis. Taking a drop-out rate of 15% in Phase B into account, 126 randomized patients per arm would be required.

In total, 378 randomized patients were required in the final analysis in addition to 68 patients was randomised to the fluoxetine group (Study part 1), which means a total of 446 randomised patients.

Per US FDA requirement, and as stipulated in the study protocol, a blinded sample size re-assessment was performed prior to the interim analysis to ensure that the study is adequately powered.

The SD estimate that appears to stabilize after approximately 160 patients (see Interim SAP) have been included. The mean value of the SDs observed after n=160 is 12.67 (with SD_{SD} = 0.1). The originally assumed SD for the study was 11. The observed withdrawal rate (WD) among the 271 patients was found to be 9.2 % since 246 patients provide data at Week 8. The originally assumed WD was 15%, thus, the observed withdrawal rate is less than originally assumed. With the updated SD and WD, the sample size of 480 patients for the fixed design was found to be insufficient to meet the target power. For a fixed design to have at least 85% power with and SD =12.67 and a withdrawal rate of 9.2%, n=150 patients would need to be randomised per group.

In the sequential approach where an interim analysis were performed for 271 randomised patients (n=68) per group, the overall sample size was adjusted to n=157 per group. This gives a total sample size of 3*157+68 = 539 to be randomised as the decision at interim was to continue the study.

16 Statistical Software

The statistical software used will be SAS® Version 9.4 or later.

17 Changes to Analyses Specified in the Protocol and protocol amendments

The following changes have been made as compared to the latest version of the protocol:

- C-SSRS categorisation has been updated and will no longer include C-CASA
- CDRS-R sub scores Mood, Somatic, Subjective, and Behaviour have been added as secondary endpoints
- CDRS-R single items have been added as an exploratory endpoint
- response in the CGI-I score (defined as a CGI-I score of 1 or 2) during the 8 week treatment period has been deleted

All previous major changes including interim analysis, average dose of vortioxetine versus placebo, and sample size re-assessment have been aligned with authorities.

18 Details on Data Handling

Unless otherwise specified, the baseline value for Phase A will the latest value captured at or before Visit 2.

Unless otherwise specified, the baseline value for Phase B will be defined as the value at Visit 6.

In the TFLs and CSR, Baseline A will be named Enrolment and Baseline B will be named Randomization.

Classification of adverse events into periods is defined in section 11.1.2.3.

For other data, assessments from unscheduled laboratory and ECG assessments will be assigned to an analysis phase according to Panel 5, while withdrawal visits and unscheduled scale assessments will be assigned to the study phase where assessment was performed (Phase A or Phase B). Scheduled visits will be assigned to an analysis phase based on the visit:

- *Screen* (10 days) - Starts at the Screening Visit and continues up to, and including, Visit 2
- *Phase A* (4 weeks) - Starts after Visit 2 and continues up to, and including, Visit 6 (including the withdrawal Visit for non-randomized patients)
- *Phase B* (8 weeks) - Starts after Visit 6 and continues up to, and including, Visit 12
- *Follow-Up Period* - Starts after last Visit in Phase A (non-randomized patients) or Phase B (randomized patients)

In the TFLs and CSR, Phase A will be named Single-Blind Period, and Phase B will be named Double-Blind Period.

Panel 5 Assigning Period to Unscheduled Laboratory and ECG Assessments

	Analysis Period
Assessment date ≤Visit 2	SCREEN
Visit 2<Assessment date ≤ Visit 6	Phase A
Assessment date>Visit 6 (randomized patients)	Phase B

18.1 Derived Variables

18.1.1 Children depression rating scale revised version (CDRS-R)

The CDRS-R is a clinician-rated scale to measure the severity of depression of children and adolescents. The CDRS-R consists of 17 items out of which 3 items rate nonverbal observations (listless speech, hypoactivity, and depressed affect). Fourteen items are rated on a 7-point scale from 1 to 7 and three items (sleep disturbance, appetite disturbance, and listless speech) are scored on a 5-point scale from 1 to 5. The total score ranges from 17

(normal) to 113 (severe depression). The CDRS-R can be administered by a clinician after a training session. It takes approximately 20 to 30 minutes to administer and rate the CDRS-R.

Four subscores[11] are defined based on the CDRS-R:

- Mood: sum of items 8, 11, 14, 15; score range 4 to 28
- Somatic: sum of items 4, 5, 6, 7, 16, 17; score ranges from 6 to 36
- Subjective: sum of items 9, 10, 12, 13; score ranges from 4 to 28
- Behaviour: sum of items 1, 2, 3; score ranges from 3 to 21

18.1.2 Clinical Global Impression scales (CGI-S/I)

The CGI was developed to provide global measures of the severity of a patient's clinical condition and improvement or worsening during clinical studies.

The clinician will use the scale to assess depressive symptoms.

The CGI consists of two clinician-rated subscales: severity of illness (CGI-S) and global improvement (CGI-I).

The CGI-S provides the clinician's impression of the patient's current state of mental illness. The clinician uses his or her clinical experience of this patient population to rate the severity of the patient's current mental illness on a 7-point scale ranging from 1 (Normal - not at all ill) to 7 (among the most extremely ill patients).

The CGI-I provides the clinician's impression of the patient's improvement (or worsening). The clinician assesses the patient's condition relative to baseline A on a 7-point scale ranging from 1 (very much improved) to 7 (very much worse). In all cases, the assessment should be made independent of whether the rater believes the improvement is drug-related or not.

An experienced clinician can use the CGI after a short training session. It takes 1 to 2 minutes to score the CGI after a clinical interview.

18.1.3 Children's Global Assessment Scale (CGAS)

The CGAS[12] is a clinician-rated global scale to measure the lowest level of functioning for a child (4 to 16 years) during a specified time period. The CGAS contains behaviourally-oriented descriptors at each anchor point that depict behaviours and life situations applicable to a child. The items range in value from 1 (most functionally impaired child) to 100 (the healthiest). A score above 70 indicates normal function. The CGAS can be administered by a clinician after a training session. It takes approximately 2 minutes to administer CGAS.

18.1.4 General Behavior Inventory (GBI) 10-Item Depression Scale

The GBI 10-item depression scale[13] is a parent- and subject-rated scale designed to screen for depressive symptoms in children and adolescents. The ten depression items are rated on a 4-point scale from 0 (never or hardly ever) to 3 (very often or almost constantly). The total

score ranges from 0 to 30, with higher scores indicating greater pathology. It takes approximately 5 minutes to complete the GBI 10-item depression scale.

18.1.5 Parent Global Assessment - Global Improvement (PGA)

The PGA[14] is a parent- rated variation of the CGI-I to evaluate the severity of the child's symptoms. The PGA reflects assessments of change from baseline A symptoms using a 7 point scale ranging from 1 (very much improved) to 7 (very much worse). It takes 1 to 2 minutes to score the PGA.

18.1.6 The PedsQL Present Functioning Visual Analogue Scales (PedsQL™ VAS)

The PedsQL™ VAS[15] is designed to measure at-that-moment functioning in children and adolescents. The PedsQL VAS consists of 6 domains: anxiety, sadness, anger, worry, fatigue and pain using visual analogue scales. The functionality for each domain is measured on a 10 cm line with a happy face at one end and a sad face at the other. The patients are asked to mark on the line how they feel. The total score is the average of all 6 items, and the emotional distress summary score is the mean of the anxiety, sadness, anger, and worry items. It takes 2 to 5 minutes to administer and rate the PedsQL VAS.

18.1.7 Paediatric Quality of Life Enjoyment and Satisfaction Questionnaire (PQ-LES-Q)

The PQ-LES-Q[16] is a patient-rated scale designed to assess satisfaction with life. It is an adaptation of the Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q), which is used to measure quality of life in adults. The PQ LES Q consist of 15 items, item 1-14 assess the degree of satisfaction experienced by subjects in various areas of daily functioning and item 15 allows subjects to summarise their experience in a global rating (overall evaluation). Each item is rated on a 5-point scale from 1 *(very poor)* to 5 *(very good).* The total score range of item 1-14 is 14 to 70, with higher scores indicating greater satisfaction. It takes 5 to 10 minutes to complete the scale.

18.1.8 Multidimensional Anxiety Scale for Children short version (MASC-10)

The MASC-10[17] is designed to screen for anxiety in children, and is based on the 39-item MASC. The MASC-10 consists of 10 items assessing physiological symptoms, social anxiety, harm avoidance, and separation/panic. The anxiety symptoms are rated on a 4-point scale from 0 (never true about me) to 3 (often true about me). Total score ranges from 0 to 30, with higher scores indicating higher levels of anxiety. The MASC-10 takes approximately 5 minutes to administer and score.

18.1.9 Tanner evaluation

Tanner staging is a scale for assessing physical development and sexual maturity during onset and progress of puberty.

The scale includes five stages of pubertal changes (called Tanner stages) separate for males and for females.

For female the 5 stages of maturation are recognized by assessing pubic hair and breast development. For male the 5 stages of maturation are recognized by assessing pubic hair, growth of penis and testicles. Post-puberty is defined as a Tanner stage = 5 in both of observed criteria.

The evaluation of Tanner stage will be performed by physician or trained nurse - they will be provided with figures depicting the somatic changes and tables describe these changes in words to facilitate the staging.

18.1.10 Paediatric Adverse Event Rating Scale (PAERS)

The PAERS[18] is a clinician-rated scale designed and validated to assess adverse events occurring in paediatric patients who are treated with psychotropic medication in clinical studies.

The PAERS consists of 45 items: 43 specific signs and symptoms and two to be specified. Each item specifies if the adverse event was present recently, if it was resolved (No, Yes), related to drug (No, Study drug, Other drug, Drug-drug interaction), its severity (Mild, Moderate, Severe, Extreme) and if it impaired function (No, Yes).

The PAERS can be administered by a clinician after a short training session. It takes approximately 10 minutes to administer and score the PAERS.

The PAERS must be applied after the non-leading, open questions on any adverse events. As this recording of adverse events should precede the use of the PAERS, the numbers of adverse events captured by these two methods potentially differ.

18.1.11 General Behavior Inventory (GBI) 10-Item Mania Scale

The GBI 10-item mania scale[19] is a parent- and subject-rated scale designed to screen for manic symptoms in children and adolescents. The 10 items are rated on a scale from 0 (never or hardly ever) to 3 (very often or almost constantly). The total score ranges from 0 to 30 points, with high scores indicating greater pathology. It takes approximately 5 minutes to complete the GBI 10-item mania scale.

18.2 Identifying rows for analyses for rating scales and safety variables

For lab, ecg, vs, and scales except PAERS and C-SSRS, assessments (scheduled, re-assessments, and unscheduled) in Phase B obtained more than 7 days after the last dose of IMP (assessment date – last dose date > 7) in the Phase or study will not be considered in descriptive statistics or statistical analyses but will be included in listings of safety variables (assessments≤7 days since last IMP will be referred to as valid). If the date of last dose is

missing for a patient in Phase B, all assessments for the patient will be considered valid. Not valid assessments will be assigned analysis Day 8888 and analysis Week 888.

For further details see Data Handling Plan.

Analysis day and week for assessments at scheduled visits (Withdrawal Visit not considered part of scheduled visits) will be the protocol (nominal) day and week:

Assessments at the Withdrawal Visit will be assigned to an analysis day in the phase from which the patient withdrew:

- Patients withdrawn in Phase A:
  Analysis day will be assigned according to Panel 6, based on days since Visit 2 (assessment date-Visit 2 date)
- Patients withdrawn from Phase B:
  Analysis day will be assigned according to Panel 7 based on days since Visit 6 (assessment date-Visit 6 date)

Panel 6 Visit Windows Patients Withdrawn from Phase A

CDRS-R PGA	GBI (depression) GBI (mania)	CGAS PedsQL VAS PQ-LES-Q MASC-10 Weight Laboratory ECG	CGI-S CGI-I PAERS Vital Signs	Analysis Phase	Target Day based on assessment date-Visit 2
			1 to 10	Phase A	7
1 to 17	1 to 21		11 to 17	Phase A	14
18 to 24			18 to 24	Phase A	21
>24	>21	>0	>24	Phase A	28

Panel 7 Visit Windows Patients Withdrawn from Phase B

CDRS-R PGA	GBI (depression) GBI (mania)	CGAS PedsQL VAS PQ-LES-Q MASC-10 Weight Laboratory ECG	CGI-S CGI-I PAERS Vital Signs	Analysis Phase	Target Day based on assessment date-Visit 6
			1 to 10	Phase B	7
1 to 21	1 to 21		11 to 17	Phase B	14
			18 to 24	Phase B	21
22 to 35	22 to 35	1 to 42	25 to 35	Phase B	28
36 to 49	36 to 49		36 to 49	Phase B	42
>49	>49	>42	>49	Phase B	56

18.3 Handling of Missing or Incomplete Dates/Times

18.3.1 IMP Start and Stop Dates by phase

In derivation of data for adverse events and concomitant medications, a missing IMP start date in Phase A or Phase B will be imputed by:

- Patients in APTS_A with a missing IMP start date in Phase A: Date of Visit 2
- Patients in APTS with a missing IMP start date in Phase B: Date of Visit 6

In evaluation of valid efficacy and safety assessments in Phase B (assessment is considered valid if assessment date-date of last IMP ≤7 days), a missing date of last IMP in Phase B will be imputed by the latest attended Visit in the study before the safety follow-up.

Exposure for patients in APTS_A/APTS with missing IMP start or stop date will not be calculated.

Exposure for patients not in APTS_A/APTS will be set to 0.

18.3.2 Medical Disorder Start and Stop Dates

Incomplete dates will not be imputed. Classification of events into *concurrent medical disorders* or *past disorders* will be based on the reported ongoing status.

18.3.3 Medication Start and Stop Dates

In order to assign all medications to an analysis start-and stop phase, missing or incomplete start-or stop dates will be imputed by phase. Ongoing medications at the end of a study will not be assigned a stop phase in that study.

No duration will be calculated for medications with imputed start-or stop date, or for ongoing medications.

For further details see Data Handling Plan.

18.3.4 Adverse Event Start Dates

In order to classify the treatment emergent status and assign all adverse event records to an analysis, missing or incomplete start-or change in intensity dates will be imputed by phase.

Incomplete adverse start dates will be imputed before handling of incomplete dates for change in intensity-or seriousness.

Please see Data Handling Plan for further details.

Incomplete adverse event stop dates will not be imputed.

18.4 Grouping of Small Countries

Countries with few patients in the FAS will be grouped for potential use in case of the MMRM not converging. A country is considered small if it contained <4 patients in the FAS.

The final grouping of countries will be done in a blinded manner at the classification meeting and will be documented in the classification meeting minutes.

References

1. World Medical Association (WMA). Declaration of Helsinki: Ethical principles for medical research involving human subjects. [Internet] wma.net/en/30publications/10policies/b3/index.html
2. Jennison C, Turnbull BW. Group Sequential Methods with Applications to Clinical Trials. Boca Raton, Florida: Chapman & Hall/CRC; 2000.
3. Carpenter JR, Kenward M. Missing data in randomized controlled trials – A practical guide. 21 November 2007.
4. Shen S, Beunckens C, Mallinckrodt CH, Molenberghs G. A local influence sensitivity analysis for incomplete longitudinal depression data. J Biopharm Stat. 2006; 16: 365-384.
5. Siddiqui O, Hung HMJ, O'Neill R. MMRM vs. LOCF: A comprehensive comparison based on simulation study and 25 NDA datasets. J Biopharm Stat. 2009; 19: 227-246.
6. Committee for Proprietary Medicinal Products (CPMP). Points to consider on adjustment for baseline covariates. 2003
7. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. Drug Information Journal. 2008; 42: 303-319
8. Little R, Yau L. Intent-to-treat analysis for longitudinal studies with drop-outs. Biometrics. 1996; 52: 1324-1333.
9. Rubin, DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1997.
10. United States Food and Drug Administration (FDA). Guidance for Industry: Drug-induced liver injury: Premarketing Clinical Evaluation. July 2009.
11. Poznanski EO, Grossman JA, Buchsbaum Y, Banegas M, Freeman L, Gibbons R. Preliminary studies of the reliability and validity of the children's depression rating scale. Journal of the American Academy of Child Psychiatry 1984; 23, 2:191-197.
12. Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, et al. A children's global assessment scale (CGAS). Arch Gen Psychiatry 1983; 40: 1228-1231.
13. Youngstrom EA, Zhao J, Mankoski R., Forbes RA, Marcus RM, Carson W, et al. Clinical significance of treatment effects with aripiprazole versus placebo in a study of manic or mixed episodes associated with pediatric bipolar I disorder Journal of Child & Adolescent Psychopharmacology 2013; 23: 72-79.
14. Biederman J, Lopez FA, Boellner SW, Chandler MC. A randomized, double-blind, placebo-controlled, parallel-group study of SLI381 (Adderall XR) in children with attention deficit/ hyperactivity disorder. Pediatrics 2002; 110: 258-266.
15. Sherman SA, Eisen S, Burwinkle TM, Varni JW. The PedsQL™ present functioning visual analogue scales: preliminary reliability and validity. Health and Quality of Life Outcomes 2006; 4: 75.

16. Endicott J, Nee J, Yang R, Wohlberg C. Pediatric Quality of Life Enjoyment and Satisfaction Questionnaire (PQ-LES-Q): reliability and validity. J Am Acad Child Adolesc Psychiatry 2006; 45: 401-407.
17. March JS, Sullivan K, Parker J. Test-retest reliability of the multidimensional anxiety scale for children. Journal of Anxiety Disorders 1999; 13: 349-358.
18. March J, Karayal O, Chrisman A. CAPTN: The Pediatric Adverse Event Rating Scale. Paper presented at: Scientific Proceedings of the 2007 Annual Meeting of the American Academy of Child and Adolescent Psychiatry; October 23 to 28, 2007; Edited by Novins DK, DeYoung. Boston, MA; 2007; 241.
19. Youngstrom EA, Frazier TW, Demeter C, Calabrese JR, Findling RL. Developing a 10-item mania scale from the Parent General Behavior Inventory for children and adolescents. J Clin Psychiatry 2008; 69: 831–839.

Appendix I
Statistical Analysis Plan
Authentication and Authorisation

Statistical Analysis Plan
Authentication and Authorisation

Study title:	Interventional, randomised, double-blind, placebo-controlled, active reference (fluoxetine), fixed-dose study of vortioxetine in paediatric patients aged 7 to 11 years, with Major depressive disorder (MDD)
SAP date:	8 February 2022

This document has been signed electronically. The signatories are listed below.

Authentication

[redacted]

Authorisation

[redacted]

Appendix II
Study Flow Chart

Study Flow Chart

		Phase A					Phase B						
		Single-blind Treatment Period					Double-blind Treatment Period						
Visit	**Screening**	**Baseline (Phase A)**				**Baseline (Phase B)**						**Completion /Withdrawal[a]**	**Safety Follow-up[b]**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Day/End of Week	**d-5/d-15**	**0**	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W10**	**W12**	**W16**
Visit Window[c] (days relative to nominal visit)			**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**
Screening/Baseline Procedures and Assessments													
Signed informed assents/consents	√												
Diagnosis (DSM-5™)	√												
K-SADS-PL[d]	√												
Disease-specific history	√												
Relevant history (social, medical, psychiatric, neurological)	√												
History of stimulant medication, if any	√												
Demographics (age, sex, race)	√												
Smoking, alcohol consumption	√												
Family psychiatric history	√												
Traumatic life events	√												
Pregnancy test[e]	√											√[f]	
Inclusion/exclusion criteria	√	√											
Randomisation to Phase B						√							
BPI			√	√	√		√			√			
Efficacy Assessments													
CDRS-R	√	√		√	√	√		√		√	√	√	
CGI-S	√	√	√	√	√	√	√	√	√	√	√	√	
CGI-I			√	√	√	√	√	√	√	√	√	√	
PGA (PRO)[g]				√	√	√		√		√	√	√	
GBI (depression subscale)	√	√		√		√		√		√	√	√[f]	
CGAS		√				√				√		√[f]	
PedsQL VAS (PRO)		√				√				√		√[f]	
PQ-LES-Q (PRO)		√				√				√		√[f]	

		Phase A					Phase B						
		Single-blind Treatment Period					Double-blind Treatment Period						
Visit	**Screening**	**Baseline (Phase A)**				**Baseline (Phase B)**						**Completion /Withdrawal[a]**	**Safety Follow-up[b]**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Day/End of Week	**d-5/d-15**	**0**	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W10**	**W12**	**W16**
Visit Window[c] (days relative to nominal visit)			**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**
Exploratory Assessments													
MASC-10 (PRO)		√				√				√		√f	
Vortioxetine and Fluoxetine Quantification													
Blood sampling vortioxetine and fluoxetine quantification										√		√f	
Translational Medicine Assessments[h]													
Blood sampling for gene expression profiling						√				√		√f	
Blood sampling for metabolomics/proteomics						√				√		√f	
Blood sampling for pharmacogenetics (optional)	√												
Safety Assessments													
Adverse events[i]	√	√	√	√	√	√	√	√	√	√	√	√	√j
Blood and urine sampling for clinical safety laboratory tests	√					√				√		√f	
Vital signs	√	√	√	√	√	√	√	√	√	√	√	√	
ECGs	√					√				√		√f	
Weight and height	√					√				√		√f	
Examinations (physical and neurological)	√											√f	
Tanner score	√												
PAERS	√	√	√	√	√	√	√	√	√	√	√	√f	
C-SSRS	√	√	√	√	√	√	√	√	√	√	√	√	
GBI (mania subscale)	√	√		√		√		√		√	√	√f	
Other Study Procedures													
IMP dispensed		√	√	√	√	√	√	√	√	√	√		
Possible change in IMP dose										√k			
IMP returned and IMP accountability			√	√	√	√	√	√	√	√	√	√	

		Phase A					Phase B						
		Single-blind Treatment Period					Double-blind Treatment Period						
Visit	**Screening**	**Baseline (Phase A)**				**Baseline (Phase B)**						**Completion /Withdrawal[a]**	**Safety Follow-up[b]**
Visit Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Day/End of Week	**d-5/d-15**	**0**	**W1**	**W2**	**W3**	**W4**	**W5**	**W6**	**W7**	**W8**	**W10**	**W12**	**W16**
Visit Window[c] (days relative to nominal visit)			**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**	**± 2**
Recent and concomitant medication	√	√	√	√	√	√	√	√	√	√	√	√	

a. This visit should take place as soon as possible after the patient withdraws from the study. Any visit during Phase A, can be converted in to a withdrawal visit on the day of the visit. The BPI is optional if a visit during Phase A is converted into a withdrawal visit.
b. This can be a telephone contact, unless an SAE has occurred since the last visit or there was a clinically significant abnormal safety laboratory test value at the last visit. In such cases, safety follow-up(s) must be scheduled to allow for a medical examination and/or blood sampling. Further safety follow-up visits beyond 30 days may be needed as judged by the investigator (if further safety follow-up visits are performed, these must be recorded in the patient's medical record, and not in the eCRF). All enrolled patients should have a safety follow-up visit unless they continue into the extension study.
c. If the date of a patient visit does not conform to the study plan, subsequent visits should be planned to maintain the visit schedule relative to randomisation/baselines.
d. The Kiddie-Schedule for Affective Disorders and Schizophrenia for School-aged Children, Present and Lifetime version (K-SADS-PL) will be used to confirmed diagnosis of Major depressive disorder (MDD) and to assess possible psychiatric co-morbidities.
e. If the patient is a female subject of childbearing potential (defined as females aged ≥10 year's old and younger girls who, at the discretion of the investigator, were deemed to be of reproductive potential), a urine pregnancy test is to be performed at screening and if the patient has completed or withdrawn in Phase B.
f. Should only be performed if the patient has completed or withdrawn in Phase B.
g. The PGA assessment should be in reference to Baseline A.
h. Samples will be collected at the designated time points. Any results of the Translational Medicine Assessments will not be available at the time of reporting and will consequently not be part of the actual report of the study.
i. Pre-treatment adverse events (before IMP intake i.e. before Baseline B) must be recorded on the *Adverse Event Form* in the eCRF. Adverse events from open questioning in addition to all spontaneously reported adverse events must be recorded on the *Adverse Event Form* in the eCRF. If the adverse event is an SAE, the *Serious Adverse Event Report Form* must also be completed. For practical reasons, questioning for adverse events could follow vital sign assessments. In addition, the tolerability will also be assessed using the Paediatric Adverse Event Rating Scale (PAERS). Application of this scale should follow after the open, non-leading question for any adverse events.
j. Only for adverse events on-going at Completion/Withdrawal and new SAEs.
k. The vortioxetine dose may be down-titrated with 5mg/day based on poor tolerability, only one down-titration is allowed. No up-titration will be allowed. The fluoxetine dose may be down-titrated to 10 mg/day based on poor tolerability only one down-titration is allowed (Only applicable during Study part 1). No up-titration will be allowed.

Appendix III
Testing Strategy

Testing strategy

The reason why both tests in step 2 can be performed at full significance level without further multiplicity adjustment follows from a formal closed testing argument.

A formal descriptions of the closed testing principle is as follows:
"*Suppose there are k hypotheses H_1,..., H_k to be tested and the overall type I error rate is α. The closed testing principle allows the rejection of any one of these elementary hypotheses, say H_i, if all possible intersection hypotheses involving H_i can be rejected by using valid local level α tests. It controls the familywise error rate for all the k hypotheses at level α in the strong sense.*"

In this study, there are 3 hypotheses that we want to multiplicity control for:

- H1: $\theta_L=0$ (Low dose is in-efficacious)
- H2: $\theta_H=0$ (High dose is in-efficacious)
- H3: $0.5* \theta_L + 0.5* \theta_H =0$ (The mean of the high and low dose is in-efficacious)

As described above Step 1 is to test H3 at full allocated alpha level. If this is rejected the procedure goes to Step 2 testing H1 and H2 also at full allocated alpha level. Following the closed testing principle this (Step 2) would require a test of the intersection hypothesis first.

The intersection hypothesis H4 that needs to be tested is:

- H4: $H1 \cap H2: \theta_L =0 \wedge \theta_H =0$ (Both doses are in-efficacious)

However, due to the testing strategy, it can be shown that H4 has already been tested: Rejecting H3 in Step 1, gives that the average of the 2 comparisons is different from 0. Therefore, it is impossible for both comparisons θ_L and θ_H to be be zero simultaneously which means that H4 is rejected. So, rejection of H3 automatically leads to rejection of the intersection hypothesis H4, mathematically since it is contained in H3, $H4 \subseteq H3$. The closed testing requirement concerning test of intersection hypotheses is therefore fulfilled and it is valid to test both H1 and H2 separately at full allocated alpha level.

Appendix IV
SAS® Code

SAS® Code

Implementation of MMRM for the primary analysis of the primary endpoint at the final stage

Interim analysis:

```
proc seqdesign
  altref=-4.0
  pss
  maxinfo=0.5865366
  boundaryscale= stdz;
  TwoSidedErrorSpending: design
  nstages=2
  method=errfuncpow(rho=2)
  boundarykey=alpha
  alt=lower
  stop=both(betaboundary=nonbinding)
  alpha=0.025
  info=cum(204 471 );
  samplesize
  model=twosamplemean(stddev= 12.67 weight=0.66666667 0.33333333);

  ods output Boundary=BOUNDARY(label="Interim: SEQDESIGN info used in
SEQTEST");
run;

ods output LSMeans=CHG  MMRM;
ods output lsmestimates=CHG  MMRM_AvgVOR(drop=StmtNo);
ods output diffs=CHG  MMRM  Diff;
ods output estimates=CHG_MMRM_Diff_AvgVOR;


proc mixed data= adcdrsr  int noclprint;
      class USUBJID COUNTRY TRT02P NOMWEEK;
      model  CHG = COUNTRY TRT02P*NOMWEEK BASE*NOMWEEK /s ddfm=kr;
      repeated NOMWEEK /subject=USUBJID type=un;
      lsmeans TRT02P*NOMWEEK/diffs;

      lsmestimate trt02p*NOMWEEK 'Average effect of VOR at Week 8'
                  [ 1,3 4] [ 1,4 4] /divisor=2;

      estimate 'Average VOR vs PBO at Week 8'
      TRT02P*NOMWEEK 0 0 0 0
            0 0 0 -1
```

```
          0 0 0 0.5
          0 0 0 0.5;

title2 "Primary analysis, including estimation of Average VOR effect and difference to PBO
(FAS, MMRM)";
run;


*************************************************************;
%**START:                                                    ;
%**Prepare and print data with results from the MMRM-analysis ;
%**Data used for printing are saved in library Interim       ;
%*************************************************************;

data CHG_MMRM(label="Interim: CHG CDRS-R Total Score (MMRM)");
length trt02p $20.;
     set CHG  MMRM_AvgVOR(in=vor) CHG_MMRM;
     if vor then do;
          TRT02P='Avg VOR';
          TRT02PN=2;
          NOMWEEK=8;
     end;
     if TRT02P='FLU' then TRT02PN=1;
     else if TRT02P='PBO' then TRT02PN=3;
     else if TRT02P='VOR10' then TRT02PN=4;
     if TRT02P='VOR20' then TRT02PN=5;
run;
proc sort data=CHG_MMRM;
     by TRT02PN TRT02P NOMWEEK;
run;

data CHG  MMRM  Diff(drop=_NOMWEEK _TRT02P Effect label="Interim: CHG CDRS-
R Total Diff (MMRM)");
length trt02p $20.;
     set CHG  MMRM  Diff  AvgVOR(in=vor)
CHG  MMRM  Diff(where=(((nomweek=_nomweek)) and (TRT02P="PBO" or
 TRT02P="PBO")));
     if vor then do;
          TRT02P='Avg VOR';
          NOMWEEK=8;
          TRT02PN=2;
     end;
     if TRT02P='PBO' then do;
          estimate=estimate*-1;
          TRT02P=_TRT02P;
     end;
```

```
    if TRT02P='FLU' then TRT02PN=1;
    else if TRT02P='PBO' then TRT02PN=3;
    else if TRT02P='VOR10' then TRT02PN=4;
    if TRT02P='VOR20' then TRT02PN=5;
run;

proc sort data=CHG  MMRM  Diff;
    by TRT02PN TRT02P NOMWEEK;
run;

title2 "Mean Change from Baseline B in CDRS-R Total Score at Week 8 (FAS, MMRM)";
proc print data=CHG  MMRM noobs;
    where NOMWEEK=8;
    var TRT02P NOMWEEK Estimate StdErr DF Tvalue probt;
run;
title2 "Mean Difference in Change from Baseline B CDRS-R Total Score Compared to PBO
at Week 8 (FAS, MMRM)";
proc print data=CHG  MMRM_Diff noobs;
    where NOMWEEK=8;
    var TRT02P NOMWEEK Estimate StdErr DF Tvalue probt;
run;

title2;
%*********************************************************************;
%**END:                                                               ;
%**Prepare and print data with results from the MMRM-analysis        ;
%*********************************************************************;


%*********************************************************************
******;
%**Prepare data for Interim Analysis using SEQTEST (required variables added)   ;
%*********************************************************************
******;

data CHG  MMRM  Diff  AvgVOR(label="Interim: Data used in SEQTEST");
    set CHG  MMRM_Diff_AvgVOR;
    _Scale_ = "MLE";
     Stage  = 1;
    variable="Treat";
    keep _Scale_ _Stage_ Estimate StdErr probt variable;
run;


%*********************************************************************
************************;
```

```
%** Run SEQTEST and print outcome (Action)
                    ;
%** Note, input Z-statistic = Estimate/STDerr from nterim.CHG_MMRM_Diff becomes 'Es-
timate' in output;
%** Input are the derived boundaries generated by SEQDESIGN
                    ;
%*************************************************************************
************************;

ods graphics on;
proc seqtest Boundary=BOUNDARY
 Parms(Testvar=treat)=CHG  MMRM_Diff_AvgVOR
 boundaryadj=errfuncpow(rho=2)
 boundarykey=alpha
 infoadj=prop;
 ods output Test=InterimTest(label="Interim: Outcome of SEQTEST");
 title2 "Analysis of Estimated Average VOR Versus Boundaries: Based on Z-statistic";
run;

title2 "Action";
footnote "Action: Continue=Study continue, Reject Null=Stop for Efficacy (Avg VOR supe-
rior to PBO), Accept Null=Stop for Futility (Avg VOR has no chance of reaching signifi-
cance)";

proc print data=InterimTest;
run;

title;
footnote;

ods graphics off;



final analyses:


ods output LSMeans=CHG_MMRM;
ods output lsmestimates=CHG  MMRM_AvgVOR(drop=StmtNo);
ods output diffs=CHG  MMRM  Diff;
ods output estimates=CHG_MMRM_Diff_AvgVOR_fin;

proc mixed data=adcdrsr  fin noclprint;
      class USUBJID COUNTRY(ref="USA") TRT02P NOMWEEK;
      model  CHG = COUNTRY TRT02P*NOMWEEK BASE*NOMWEEK /s ddfm=kr;
      repeated NOMWEEK /subject=USUBJID type=un;
```

```
    lsmeans TRT02P*NOMWEEK/diffs;

    lsmestimate trt02p*NOMWEEK 'Average effect of VOR at Week 8'
            [ 1,3 4] [ 1,4 4] /divisor=2;

    estimate 'Average VOR vs PBO at Week 8'
    TRT02P*NOMWEEK 0 0 0 0
        0 0 0 -1
        0 0 0 0.5
        0 0 0 0.5;


title2 "Primary analysis, including estimation of Average VOR effect and difference to PBO
(FAS, MMRM)";
run;

data finaltestdata;
  set CHG  MMRM_Diff_AvgVOR_fin;
    Scale   = 'MLE';
    Stage  = 2 ;
  variable="Treat";
  keep _Scale_ _Stage_ Estimate StdErr probt variable tValue;
run;

ods graphics on;
proc seqtest Boundary=InterimTest
    Parms(Testvar=treat)=finaltestdata
    boundaryadj=errfuncpow(rho=2)
    boundarykey=alpha
    boundaryscale= stdz
    errspend
    infoadj=prop
    cialpha=0.05
    citype=twosided
    order=stagewise;
    ods output Test=finTest
        Design=finDesign;
run;
ods graphics off;
```

Implementation of MMRM for analysing continuous endpoints

```
proc mixed noclprint data=xyz ic method=reml;

class usubjid country trt02p  nomweek;

   model chg = base country nomweek  trt02p  trt02p * nomweek base* nomweek /s ddfm=kr;

   repeated nomweek /subject=usubjid type=un;

   lsmeans trt02p * nomweek/diff;

run;
```

Implementation of ANCOVA for analysing continuous endpoints

```
proc mixed data=xyz method=reml;
   class country trt02p  ;
   model chg=base country trt02p  / ddfm=kr;
   lsmeans actarm / diff=control("Placebo") cl alpha=0.05;
run;
```

Implementation of logistic regression for analysing dichotomous endpoints

```
proc logistic data=xyz;
   class trt02p / param=GLM;
   model y (event="1") = trt02p base/ clodds=wald alpha=0.05;
run;
```

Implementation of MNAR Sensitivity analysis for primary endpoint

The SAS code fragments described in this section are expected to form the basis of the MNAR sensitivity analyses. Below, we assume that:

- anadat: the analysis dataset
- anadat_mono: the analysis dataset with non-monotone missing data being imputed using MCMC
- score_5 to score_12: post-baseline CDRS-R total scores
- lastvis: the last visit at which data are observed

Imputation of non-monotone missing patterns using MCMC:

```
PROC SORT DATA=anadat;
    BY trt02p;
RUN;

PROC MI DATA=anadat OUT=anadat_mono nimpute=1000 SEED=xxxxx;
    BY trt02p;
    VAR country base score_5-score_12;
    MCMC CHAIN=MULTIPLE IMPUTE=MONOTONE;
RUN;
```

The following SAS code will be used for the PMI at the ith visit, assuming the dataset impd_&i has been imputed before the ith visit. Once the missing data are imputed sequentially, the MMRM analysis can be performed and their results can be combined using PROC MIANALYZE.

```
DATA anadat_PMI;
    SET anadat_mono;
    WHERE trt02p ="PBO";
RUN;

DATA imp_&i out_&i;
    SET impd_&i;
    IF trt02p ne “PBO” AND lastvis >= &i
    THEN OUTPUT out_&i;
    ELSE OUTPUT imp_&i;
RUN;

PROC MI DATA=imp_&i OUT=imptd_&i NIMP=1 SEED=xxxx;
    BY _imputation_;
    VAR country base score_5-score_&i;
    MONOTONE REG (score_&i = country base %DO _v=5 %TO %eval(&i-1);
score_&v. %END;);
RUN;

DATA impd_%eval(&i+1);
    SET out_&i imptd_&i;
RUN;
```

Assuming datain is the dataset with missing data imputed the analysis will proceed with:

```
PROC MIXED DATA= datain IC METHOD=REML;
    BY _Imputation_;
    CLASS usubjid trt02p country nomweek;
    MODEL chg= country trt02p*nomweek base * nomweek /s DDFM=KR;
    REPEAT nomweek / subject=usubjid type=UN;
    LSMEAN trt02p*nomweek / diff cl alpha=0.05;
    ODS OUTPUT diffs=dataout_d;
```

```
    ODS OUTPUT lsmeans=dataout_l;
RUN;



DATA dataout_d;
    SET dataout_d;
    WHERE nomweek=_nomweek and trt02p ne "PBO" and _trt02p = "PBO";
RUN;

PROC SORT data=dataout_l OUT=dataout_l NODUPKEY;
    BY _Imputation_ trt02p nomweek;
RUN;

PROC SORT DATA=dataout_d; BY nomweek _Imputation_; run;

PROC SORT DATA=dataout_l; BY nomweek _Imputation_; run;

PROC MIANALYZE PARMS(classvar=full)=dataout_d;
    CLASS trt02p;
    MODELEFFECTS nomweek*trt02p;
    ODS OUTPUT ParameterEstimates=dataout_d1;
    BY visit;
RUN;
```

Appendix V
PCS Criteria

PCS Criteria

Table 1 Lundbeck PCS Criteria for Clinical Safety Laboratory Tests

Laboratory Test	CDISC Term	Unit	PCS LOW	PCS HIGH
Haematology / Coagulation				
B-haemoglobin	HGB	g/dL	≤ 9.5 (women); ≤ 11.5 (men)	≥ 16.5 (women); ≥ 18.5 (men)
B-erythrocytes (red cell count)	RBC	x 10E12/L	≤ 3.5 (women); ≤ 3.8 (men)	≥ 6.0 (women); ≥ 7.0 (men)
B-haematocrit (packed cell volume)	HCT	V/V	≤ 0.32 (women); ≤ 0.37 (men)	≥ 0.50 (women); ≥ 0.55 (men)
B-MCV (mean cell volume)	MCV	fL	≤ 0.8 x LLN	≥ 1.2 x ULN
B-total leucocyte (white cell count)	WBC	x 10E9/L	≤ 2.8	≥ 16
B-neutrophils/leucocytes	NEUTLE	%	≤ 20	≥ 85
B-eosinophils/leucocytes	EOSLE	%		≥ 10
B-basophils/leucocytes	BASOLE	%		≥ 10
B-lymphocytes/leucocytes	LYMLE	%	≤ 10	≥ 75
B-monocytes/leucocytes	MONOLE	%		≥ 15
B-thrombocytes (platelet count)	PLAT	x 10E9/L	≤ 75	≥ 600
P-INR (prothrombin ratio)	INR	Ratio		≥ 2.0
B-prothrombin time	PT	Sec		≥ 18
Liver				
S-aspartate aminotransferase	AST	IU/L		≥ 3 × ULN
S-alanine aminotransferase	ALT	IU/L		≥ 3 × ULN
S-bilirubin	BILI	µmol/L		≥ 34
S-bilirubin, direct	BILDIR	µmol/L		≥ 12
S-bilirubin, indirect	BILIND	µmol/L		≥ 22
S-alkaline phosphatase	ALP	IU/L		≥ 3 × ULN
S-gamma glutamyl transferase	GGT	IU/L		≥ 200
S-alpha-glutathione S-transferase (alpha-GST)	GSTAL	µg/L		≥ 20
Kidney				
S-creatinine	CREAT	µmol/L		≥ 1.5 x ULN
B-urea nitrogen (BUN)	BUN	mmol/L		≥ 11
S-uric acid (urate)	URATE	µmol/L		≥ 510 (women); ≥ 630 (men)
Electrolytes				
S-sodium (natrium)	SODIUM	mmol/L	≤ 125	≥ 155
S-potassium (kalium)	K	mmol/L	≤ 3.0	≥ 6.0
S-calcium	CA	mmol/L	≤ 1.8	≥ 3.0
S-chloride	CL MG	mmol/L	≤ 90	≥ 117
S-magnesium	PHOS	mmol/L	≤ 0.6	≥ 1.3
S-phosphate (phosphorus, (inorganic)	BICARB	mmol/L	≤ 0.65	≥ 1.95
S-bicarbonate		mmol/L	≤ 12	≥ 38

Endocrine / Metabolic				
B-glucose, non-fasting/unknown	GLUC	mmol/L	≤ 3.4	≥ 9.4
B-glucose, fasting	GLUC	mmol/L	≤ 3.0	≥ 6.0
S-glucose, non-fasting/unknown	GLUC	mmol/L	≤ 3.9	≥ 11.1
S-glucose, fasting	GLUC	mmol/L	≤ 3.5	≥ 7.0
B-glycosylated haemoglobin, fasting	HBA1C	%		≥ 6.5
S-prolactin	PROLCTN	mIU/L		≥ 1350
S-thyrotropin/TSH	TSH	mIU/L	≤ 0.3	≥ 5.5
S-protein (total)	PROT	g/L	≤ 45	≥ 95
S-albumin	ALB	g/L	≤ 27	
Lipids				
S-cholesterol total, non-fasting/unknown	CHOL	mmol/L		≥ 7.8
S-cholesterol total, fasting	CHOL	mmol/L		≥ 6.2
S-triglycerides, non-fasting/unknown	TRIG	mmol/L		≥ 5.65
S-triglycerides, fasting	TRIG	mmol/L		≥ 4.2
S-LDL cholesterol, non-fasting/unknown	LDL	mmol/L		≥ 5.3
S-LDL cholesterol, fasting	LDL	mmol/L		≥ 4.9
S-HDL cholesterol, non-fasting/unknown	HDL	mmol/L	≤ 0.8	
S-HDL cholesterol, fasting	HDL	mmol/L	≤ 0.9	
Cardiac / Skeletal/Muscle				
S-creatine kinase (total)	CK	IU/L		≥ 400 (women); ≥ 750 (men)
S-creatine kinase MB isoenzyme	CKMB	µg/L		≥ 8.5 or
	CKMBCK	%		≥ 3.5% of total CK
S-lactate dehydrogenase (total)	LDH	IU/L		≥ 750
S-troponin I	TROPONI	µg/L		≥ 1.5
S-troponin T	TROPONT	µg/L		≥ 0.4
Infection				
S-C-reactive protein	CRP	mg/L		≥ 25
S-globulin (total)	GLOBUL	g/L	≤ 15	≥ 55
Urine				
Urinary pH	PH		≤ 4	≥ 9

S=serum; B=whole blood; U=urine

Table 2 Lundbeck PCS Criteria for ECG Parameters

ECG Parameter	CDISC Term	Unit	PCS LOW	PCS HIGH
Absolute Time Interval				
PR interval	PRMEAN	Msec		≥ 260
QRS interval	QRSDUR	Msec		≥ 150
QT interval	QTMEAN	Msec		≥ 500
Derived Time Interval				
Heart rate	HRMEAN	beats/min	< 50 and decrease ≥ 15	≥ 120 and increase ≥ 15
QTcB interval	QTCB	Msec	< 300	> 500 or increase > 60
QTcF interval	QTCF	Msec	< 300	> 500 or increase > 60

Increase/decrease is relative to the baseline value

Table 3 PCS Criteria for Vital Signs, Weight/BMI, and Waist Circumference

Parameter	CDISC Term	Unit	PCS LOW	PCS HIGH
Waist circumference	WSTCIR	Cm	decrease ≥ 7%	increase ≥ 7%
Weight	WEIGHT	Kg	decrease ≥ 7%	increase ≥ 7%
Body Mass Index	BMI	kg/m2	decrease ≥ 7%	increase ≥ 7%
Pulse rate, supine/sitting/unknown	PULSE	beats/min	< 50 and decrease ≥ 15	≥ 120 and increase ≥ 15
Diastolic blood pressure, sitting/unknown	DIABP	mmHg	≤ 50 and decrease ≥ 15	≥ 105 and increase ≥ 15
Systolic blood pressure, sitting/unknown	SYSBP	mmHg	≤ 90 and decrease ≥ 20	≥ 180 and increase ≥ 20
Orthostatic systolic blood pressure[a]	OBP	mmHg	≤ -30	
Orthostatic pulse rate[a]	OPR	beats/min		≥ 20
Temperature[b]	TEMP	°C	decrease ≥ 2	≥ 38 3 and increase ≥ 2

Increase/decrease is relative to the baseline value

[a] For definition of orthostatic blood pressure and pulse rate, see text after Table 3

[b] Note, the diurnal variation may affect the temperature. Morning temperature is lower

Table 4 Normal ranges and PCS Criteria

	Normal Range (mm Hg)	PCS High (mm Hg)	PCS Low (mm Hg)
Systolic blood pressure, supine	7-9 years: 80-115 10-12 years: 85-120 13-17 years: 90-130 ≥ 18 years: 110-140	≥ 140 and increase ≥ 20 ≥ 140 and increase ≥ 20 ≥ 140 and increase ≥ 20 ≥ 180 and increase ≥ 20	≤ 90 and decrease ≥ 20
Diastolic blood pressure, supine	7-9 years: 40-75 10-12 years: 40-80 13-17 years: 40-80 ≥ 18 years: 51-85	≥ 90 and increase ≥ 15 ≥ 90 and increase ≥ 15 ≥ 90 and increase ≥ 15 ≥ 105 and increase ≥ 15	≤ 50 and decrease ≥ 15